图书情报工作
第66卷 第20期 2022年10月

引文分析视角下的中国古诗词用典问题初探

■ 蔡竹娟 赵丹群
北京大学信息管理系 北京 100871

摘 要：[目的/意义]基于引文分析视角对中国古诗词用典问题进行理论分析和实证研究，致力于促进传统人文研究范式向当前大数据时代基于数据驱动的量化研究范式的转变与拓展，推动中国古诗词用典研究与文献计量学方法的交叉融合与创新性发展。[方法/过程]首先通过对中国古诗词用典行为(现象)与学术论文引用行为(现象)之间异同点的比较对照，对古诗词用典现象的产生与发展、用典动机与用典行为、用典关系与用典网络等进行理论分析；其次，选用《全宋词》和《全宋词典故辞典》等语料，从施典维度和典源维度对宋词用典开展量化实证研究。[结果/结论]中国古诗词用典现象历史悠久，研究议题丰富，现阶段亟需研究范式的创新和转变。通过理论分析和实证结果可以看出，引文分析法有助于在较大规模古诗词语料上开展更深入的探索。目前研究存在的最大困难在于典故自动识别和提取。未来需加强与 NLP 技术的密切结合，致力在典故识别提取、用典动机与用典行为分类、用典网络与典故知识图谱构建等问题上提出更有效的自动化解决方案和量化分析策略，以有力促进传统人文研究范式由理论阐释到实证解读、由主观判断到客观揭示、由定性分析到定量分析的转变，并在更多数字人文研究议题上产生出具有创新性的成果。

关键词：引文分析 中国古诗词 典故 用典动机 用典行为
分类号：G301 I207.2
DOI：10.13266/j.issn.0252-3116.2022.20.009

1 引言

中华文明源远流长，历数千年积累、传承下来的文化典籍卷帙浩繁，古诗词无疑是其中最为璀璨夺目的一部分。而在中国古诗词的创作过程中，典故使用(用典)是常见的修辞技巧之一，历代文人墨客对此无不颇费匠心和巧思。自《诗经》始，古诗词用典现象不断绵延发展，诗词用典类型与数量、用典方式与风格偏好、典故来源及考证、典故意象及意涵嬗变等，均成为后世人文学者关注的热门话题和古诗词研究的重点内容，取得的研究成果更是十分丰硕。从研究方法上来看，古诗词用典研究主要可归结为以下3类：①传统人文研究，它通过学者个人的博闻强识和潜心钻研，着重从文学、史学、修辞学等学术视角展开相关论述；②量化研究，主要通过数学和统计学方法对某类(或特定)诗词作品及其作者的用典行为、典籍文献对古诗词创作的影响等展开定量分析；③运用自然语言处理(NLP)技术对各种诗词用典问题尝试自动化分析和算法试验。时至今日，第一类研究占据了绝大多数，形成的研究范式也处于主导和强势地位，后两类研究则出现时间较晚，相关范式尚未成形。进一步地，笔者通过文献调研发现，传统人文类研究虽然成果丰硕，涉及的研究议题也十分广泛，但大多聚焦于微观层面，以定性分析和例证说明为主，虽已成多家之言，但普遍缺少大样本数据的支持。此外，囿于学者的个人能力，研究结论不仅主观性较强，研究规模也有限，少有中观或宏观层面研究，更难以顾及或深入到所有诗词作者及所有诗词文本，越来越难以适应当前大数据时代进一步深化研究的需要。

创立于20世纪60年代的引文分析法是一种典型的“数据驱动型”研究方法，它主要致力于对学术文献(集合)中广泛存在的引用/被引用关系进行定量化研究。作为文献计量学的核心方法之一，发展至今，它已从最初的简单引文计数统计迈进到引文关联关系聚类、引文网络构建分析和知识图谱绘制(可视化)的新研究阶段，在众多领域获得了广泛应用。不难发现，一

作者简介：蔡竹娟，硕士研究生；赵丹群，教授，博士，博士生导师，通信作者，E-mail：zdq@pku.edu.cn。
收稿日期：2022-07-29 修回日期：2022-09-12 本文起止页码：82-92 本文责任编辑：徐健

蔡竹娟, 赵丹群. 引文分析视角下的中国古诗词用典问题初探[J]. 图书情报工作,2022,66(20):82 -92.

方面,古诗词用典与学术论文引用之间具有某种天然的相似性;但另一方面,两者源流不同,彼此间的差异性或自身的特殊性在所难免。有鉴于此,笔者拟将引文分析法引入到中国古诗词用典研究中,在与学术论文引用异同点进行全面对照比较的基础上,选用《全宋词》等语料对相关的古诗词用典现象展开较大规模的量化分析和实证研究。期待本研究工作及研究努力能助益于传统人文研究范式向当前大数据时代基于数据驱动的量化研究范式的转变与拓展,进而带动古诗词用典研究与文献计量学方法的深入交叉融合和创新性发展。

2 古诗词用典的理论分析

本节主要基于引文分析视角对中国古诗词用典问题展开理论分析,并尝试对古诗词用典与学术论文引用之间的异同点进行较完整的归纳总结。

2.1 用典现象的产生与演变发展

引文现象主要伴随近代科学的快速发展和大量科学发现优先权争论(纠纷)而来,学术期刊的问世以及学术界参考习惯(reference tradition)的形成,促进了引文机制的建立和完善,并使其逐渐成为从事科学研究活动需要严格遵循的基本学术规范之一。20 世纪 60 年代初期,学术文献中普遍存在的引用/被引用关系及其蕴含的检索价值、学术评价价值等被 E. Garfield 博士发现,并通过《科学引文索引》(SCI)的编制与使用得到了充分挖掘和释放。自此,引文分析法应运而生,并得到了持续研究和广泛应用。

与引文现象相比,中国古诗词用典现象的出现更为久远,最早可追溯至《诗经》[1]。作为中国古代诗歌的开端,《诗经》不仅是古诗词的用典源头,更是后世众多文学作品的典源文献(典故出处),引《诗》传统由此开启。

古诗词使用的典故主要分"事典"和"语典"两类:事典主要指古代故事,如史实、戏曲、神话、传说、典章名物等;语典主要指"有来历的现成话",如已有作品中有影响的诗、词、文佳句等[2]。概括来看,自先秦、经两汉、到魏晋南北朝,古诗词用典一直延续着引《诗》传统[3-4]。除此之外,《楚辞》《周易》《老子》《庄子》《荀子》《淮南子》《吕氏春秋》《吴越春秋》《韩诗外传》《说苑》,以及乐府、歌谣、谚语等也陆续作为典源文献被使用[5]。东晋时道家典故的增加,南朝文学"用典"手法的风行,以及诗人对典故凝练浓缩能力的增强[6]、文论家把诗歌用典当作论述对象[7]等,进一步促进了诗歌用典的发展,并在用典数量、频率、密度上都有了显著的提升。

及至唐宋,古诗词创作空前繁荣,"用典"业已成为极重要的创作习惯。自初唐起,经中唐、盛唐至晚唐,唐诗用典的取材范围极其广泛,涵盖了众多儒释道典籍及神话小说等[8]。进入宋,诗歌创作呈现出"以文字为诗、以议论为诗、以才学为诗"的特点,即宋诗的散文化、议论化以及喜用典故[9-10],用典技法更为纯熟,如黄庭坚就曾提出"夺胎换骨""点铁成金"等[11]。随着曲子词的发展,宋词创作中的用典量逐渐增多[12],而且在填词时喜欢化用唐诗,尤其是晚唐五代诗[13]。此外,王安石提倡的集句诗/词和苏轼开拓的檃括词(通过对其他诗文进行剪裁改写而成,目的主要在于"使就声律,便于歌之"[14-15]),使诗词用典(语典)发展到极致。在词人方面,南宋辛弃疾是用典大家,对稼轩词的用典方式,曾被学人总结为"正用""反用""活用""化用""暗用""连用"和"集句"7 种[16]。

宋以后的元明清三代,中国古诗词创作巅峰已过,诗词用典与唐宋时期一脉相承,少有重大变革[17]。

从以上简要回顾不难看出,中国古诗词用典现象的产生及演变发展是一个典故数量不断增多、用典频率(密度)不断加大、典故来源范围不断扩增、用典技法不断丰富的过程。笔者曾根据 Chinses - Poetry 中国诗歌古典文集数据库(下文简称"CPDB")[18]做过调研,发现宋词用典现象非常普遍,而且不仅在正文中用典,在题目和题序中亦会使用,另外还有"一典多源""一典多面""典中典""句句皆典"(集句诗/词和檃括词)等特殊而有趣的现象。

2.2 用典动机与用典行为分析

心理学家认为,动机的产生主要建立在各种不同层次需要的基础上,动机是构成人类大部分行为的基础,动机与行为之间具有紧密的关联。引文动机是一种社会性动机,主要缘于学术研究需要而产生,而引用行为可视为引文动机的一种外化表现。

引文动机(行为)问题关乎引文分析的研究假设是否成立、理论基础是否完备,因此作为一个中心性话题一直备受关注。E. Garfield 最早通过调研观察,将学术论文写作中出现的引文动机总结归纳为 15 种[19]。此后,众多学者基于不同方法和不同分析维度(框架)对其开展过大量研讨[20-23]。近年来,基于大规模语料库和 NLP 技术对引文动机(行为)进行自动识别(分类)更是成为前沿性的研究议题之一。

以下对中国古诗词用典动机与用典行为进行讨

论,并就其特殊之处加以简要说明。

2.2.1 用典动机及类型

总体说来,古诗词用典动机与引文动机一样复杂多变。作为一种文学修辞手法,作者在诗词创作中用典,或寄情表意,或说事明理,或借古讽今,或托物言志,抑或是自证博学、别出心裁等,旨趣迥异,用典的动机、目的或偏好等都极富个人色彩。不过概括来说,古诗词用典动机可粗略归为以下4类[24]:①证言,引用前人言论或事迹来证明自己的观点,或提供做事理由;②衬言,通过对典故与当下所叙之事进行比较来显示优劣、表达情感或好恶;③代名,用典面来替代当下所叙之事或物;④代言,将典故语境与诗词语境进行对照,产生言外之意,表达真正意涵。

细加考察,不难发现"证言"类动机与引文动机最为相似,其他3类则更多出于文学创作自身的需要。"证言"最能体现中国人文研究中"引经据典""无一字无来历"的治学传统。有学者对《古文观止》和《古文辞类纂》等古籍中引文现象做过调查,也发现此类动机最为普遍[25]。

正如对引文动机的识别面临着如何"走进作者脑袋"的问题,对诗词用典动机的研究也面临同样的困难。如何结合时代背景、个人际遇、诗词内容与创作情感以及典故的来源与意涵等多方面信息,运用NLP技术对古诗词用典动机展开更深入的分析和更细致的分类,其难度和挑战性或许更胜过对引文动机的研究。毕竟,学术中的引用更客观和理性,而文学中的用典则更主观和感性。

2.2.2 用典行为分析

与引用行为一样,用典行为是用典动机的外化。无论事典还是语典,一个用典行为可由以下3部分组成:①施典者(用典者),具体指诗词作者或诗词作品本身(可称"施典文献");②典故名称(又称"典面"),指所用典故在诗词中的文字表达;③典故出处,包括典源(典故最早脱胎的故事、传说或语句)和典源文献(最早对典源予以记录的典籍或文献等)[26]。

学术引用有显式引用与隐式引用之分,其中绝大多数都属于显式引用,即论文作者在文中引用他人文献时,都使用了明确且规范化的引文标识符加以标记。常用的引文标识符系统主要有顺序编码制和哈佛体系等。与之类似的是,诗词用典也有"明引"和"暗用"之分[27],其中明确告知典故来源的用典方式为"明引",否则为"暗用"。但与之形成鲜明反差的是,绝大多数诗词的用典方式都是"暗用",只有很少量的诗词会采用"明引"方式(如在诗词题序中告知典源文献、部分集句诗词会额外标注典故出处等)。如此一来,判断诗词是否用典的重要线索就主要依靠"典面"了,或者也可将"典面"理解为诗词用典的一种标识符。

不同于引文标识符的规范使用、语义明确和标注位置醒目,古诗词中出现或使用的"典面"通常富有变化,形成过程也相对复杂漫长,是"内聚性"与"外向性"相互作用的结果。这里,"内聚性"主要指典面的相对固定化,"外向性"则是指诗人/词人在用典过程中,为更好满足自身用典需求或诗词的格律平仄,会对已有典面进行调整,甚至会引入新的典面[24]。因此,一个典故可能具有多个不同的典面("一典多面")。杜甫诗句"一片花飞减却春"就是一个比较典型的"一典多面"示例。该诗句作为典面常被宋代词人直接引用,如韦骧《减字木兰花·惜春词》、米芾《鹧鸪天·漫寿》、毛滂《忆秦娥·二月二十三日夜松轩作》等,但在向滈《减字木兰花·多情被恼》、吴则礼《踏莎行·晚春》和洪咨夔《风流子·和杨帅芍》等词中,词人则对该典面(语典)进行了换字、裁字等修改操作。如此看来,以暗用为主、典面不固定("一典多面"或"异符同源")等现象的存在,为古诗词用典分析带来了很大的困扰,即使有前人诗词典故研究的大量成果可以利用(如《全宋词典故辞典》[2]《全唐诗典故辞典》[28]《诗词典故词典》[29]《中国典故大辞典》(辞书版)[30]等),典故自动识别目前也仅停留在字面相似匹配,对"一典多面"或"异符同源"等问题,仍需借助典故领域本体构建、典故同义对照工具或诗词研究专家的人工辅助[31-33]。

以下是诗词用典中的一些特殊现象,它们不仅具有很好的文学研究价值,也值得计量学学者深入研究探讨:

(1)"一典多源"。除"一典多面"外,古诗词用典还存在"一典多源"现象,即同一典故可能具有多个不同的来源或出处(典源文献),与典源文献之间可能存在着"一对多"关系。例如典故"一枝藤"的典源就有两个:一是唐·李商隐的《北青萝》;二是唐·李洞的《送人之天台》。因此,若某诗词使用了典故"一枝藤",则它与上述两个典源文献就同时建立了用典关系。

(2)"同符异源"。"同符异源"主要指不同诗词所用典故的典面相同、但典源文献却不同的现象。例如朱敦儒《聒龙谣·凭月携箫》"劝阿母、偏与金桃,教酒星、剩斟琼醑"一句中的"酒星",其典源文献为《三国

志》与《晋书》；而吴泳《水龙吟・寿李长孺》"天然带得，酒星风骨，诗囊才调"一句中的"酒星"，典源文献则为唐代诗人皮日休的《七爱诗・李翰林白》。

"同符异源"与"一典多源"容易让人混淆，两者的不同之处在于："一典多源"中只有一个典故，多个典源文献的存在并不会改变这个典故的含义；而"同符异源"则具有多个典故，且这多个典故所具有的含义并不相关(或相同)。

(3)"叠引"。"叠引"是指在一个诗词句(或句群)内连续使用多个相同或不同典故的现象[24]，与"单引"相对。古诗词中的叠引现象和叠引形式十分丰富，根据典故间的关系，大体可分为纯叠、框叠和层叠3类[24]。其中"层叠"方式最为特殊，可通俗理解为"典中典"或典故嵌套。

(4)"重复用典"。"重复用典"是一种特殊的叠引，即在一首诗词或一个诗词句(或句群)内多次引用同一典故(典面可能相同或不同)。例如辛弃疾在《贺新郎・赋滕王阁》中所写两句"王郎健笔夸翘楚。到如今、落霞孤鹜，竞传佳句"，便重复使用了同一典故，其典面为"王郎健笔"和"落霞孤鹜"，典源文献为《新唐书》卷二〇一《文艺传・王勃传》以及五代・王定保《唐摭言》卷五《以其人不称才试而后惊》。

(5)"集句用典"。"集句用典"主要指通过汇集别人现成的诗句、词句或文句来得到一首新诗或新词[34]。由于北宋王安石是第一个开始大量创作集句诗词的文人，因此后人也将之称为"荆公体"[35]。集句用典有集多家作品的，也有专集一人的，例如宋代的"集杜诗"便专集杜甫一人的诗句。在众多文人中，文天祥取得了集杜诗的最高成就[36]。

笔者调查发现，集句诗词用典逐渐开始采用了"明引"方式。例如苏轼就在他的《南乡子・怅望送春杯》中标注了所集各句的词人姓名[37]，文天祥在他的《集杜诗》诗集中则标注了每一句诗所属的杜诗标题，若标题过长，则进行缩写[38]。

历代对"集句用典"方式多有争论，有人将之视为"偷诗"，并视具体情形分为"偷语""偷意"和"偷势"等不同类型，其中"偷语最为钝贼"[39]。依此观点来看，集句诗词无疑是"偷诗之最"。但与现代学术论文写作所要求的规范引用不同，古人若想化用别人诗句且不露痕迹，抑或集出一首好诗词来，不仅要博闻强识，还需充分施展个人的创作才华。此外，"偷诗"说法还不免抹杀了集句诗词所具有的辑佚、校勘等文献学价值[40]。

2.3 古诗词用典关系与用典网络构建

古诗词用典的关系链和用典网络，可参照学术论文的引用关系和引用网络来构建，并可据此开展相关的实证研究和量化分析。

2.3.1 用典关系分析

一个基本的古诗词用典关系链条描述如下(其中方框内文字为节点，小括号内文字为关系)：

[作者(诗人/词人)] -(创作)- [诗词] -(使用/引用)- [典故] -(源自)- [典源文献]

在此关系链条中，主要存在如下两大类用典关系：

(1)不同节点(对象)之间关系。例如："作者"与"诗词"之间的"创作/被创作"关系，"诗词"与"典故"之间的"使用(引用)/被使用(被引用)"关系，"典故"与"典源文献"之间的"源自/产生"关系等。此外，由于"一典多源""同符异源""异符同源"("一典多面")等现象，典故、典面和典源文献三者之间还相互存在着多种不同的"一对多"关系。

古代学者针对这类关系，特别是围绕"典故"这一关键节点，已做出很多研究，中国古典文学中的"注典"现象即为典型代表。注家们的"注典"工作大致可分为两部分：一是指出文句的出处，二是解释语词的意义[41]。例如李善在《<文选>注》中对曹子建《七哀》诗的注释[42]，不仅建立起《七哀》诗与《古诗》《汉书》《史记》之间的用典关系，还创造了一种新的文献关系——"解释/被解释"关系。此外，"注典"是注家对诗词用典的个人解读，对典源的判断通常受限于自身的认知理解，因此不同注家对同一诗句所做解释可能会有所不同[43]。

在古人大量"注典"成果的基础上，可借鉴引文分析法对相关议题开展大规模的实证研究和可视化分析。

(2)同类节点(对象)之间关系。同类节点之间的关系主要包括以下两类：

——耦合关系。例如：若两首诗词使用了相同典故，则它们之间存在着典故耦合关系，其耦合强度可用相同典故的数量加以测度；若两首诗词所用典故的典源文献相同，则它们之间存在典源文献耦合关系，相同典源文献的数量可作为其耦合强度。

以上两种耦合关系，前者较后者限定更为严格，因为不同典故的典源文献可能相同。另外，在计算后者耦合强度时，需考虑典源文献的重复问题。例如：甲诗和乙词都使用了 n 个典故，这些典故都源于同一部典

源文献，则甲、乙的典源文献耦合强度计算有 1 和 n 两种结果，其中去重（强度为 1）代表更关注甲乙间联系的多样性，不去重（强度为 n）代表更关注甲乙间联系的紧密度。

——同被引关系（或共现关系）。例如：若两个不同典故在一首或多首诗词中被同时使用，则这两个典故之间存在同被引关系（或典故共现关系），同时使用这两个典故的诗词数量可作为其同被引强度（或共现强度）；若两部典源文献在一首或多首诗词中被同时使用，则这两部典源文献之间存在同被引关系（或典源文献共现关系），同时使用这两部典源文献的诗词数量可作为其同被引强度（或共现强度）。

值得注意的是，学术论文的同被引关系（强度）一般会随时间动态变化（单调递增），而不论是典故同被引还是典源文献同被引，两者的强度值通常都是固定不变的，这主要是因为中国古诗词与中国古代典籍都已沉淀为历史遗存，不论是诗词数量、典源文献数量，还是诗词用典情况，都已不会再增加或改变。

截至目前，基于耦合、同被引关系的古诗词用典研究及相关的实证分析都极为少见，而引文分析法对此则大有用武之地。因为针对文献（及其书目要素）之间的各种耦合和同被引关系的分析，恰是引文分析法应用的精髓所在。

2.3.2 用典网络构建

理论上，一个用典网络主要包括节点集合和节点间关系集合两大部分。根据节点间关系的不同，用典网络可分为以下两大类：

（1）同构网络。即网络中的所有节点为同一类型，节点间关系也为同一类型。例如诗词耦合网络、典故同被引网络（或共现网络）、典源文献同被引网络（或共现网络）等。视关系权重区分与否，还可细分为无权网络或权重网络。

（2）异构网络。即网络中包含多种不同节点或多种不同节点间关系，如“作者－诗词－典故”网络、“诗词－典故－典源文献”网络、“作者－诗词－典故－典源文献”网络等。事实上，此类网络也可视为一种诗词用典的“语义网”或“知识图谱”。

有关用典网络的构建及实证研究，引文分析技术目前也完全可以胜任，而这也是传统人文研究范式比较欠缺且无法触及之处。因此，从这个意义上来说，引文分析法的引入对古诗词用典问题研究是一种创新，也是在研究范式层面上的一次重要改变。

2.4 理论分析小结

以上基于引文分析视角对中国古诗词的用典问题开展了多维度的理论分析。研究发现，古诗词用典与学术引证之间既存在着天然的相似性，也表现出诸多不同，两者异同点的总结对照见表 1。

3 古诗词用典的实证研究

从古诗词用典现象的发展历史来看，唐宋时期已达巅峰，因此对古诗词用典的实证研究及分析讨论，首选以唐诗或宋词为例最为适宜。限于篇幅，以下主要使用《全宋词》和《全宋词典故辞典》等语料数据，对古诗词用典的实证研究进行设计，并据此开展相关的宋词用典实证分析。

3.1 语料选取及预处理

3.1.1 语料选取

宋词用典实证研究所需要的语料主要包括《全宋词》《全宋词典故辞典》以及《全宋词作者索引》等。

（1）《全宋词》（主语料）。自宋已降，历代皆有对宋人词集的编纂整理，唐圭璋先生编纂的《全宋词》[45]无疑是其中的集大成之作，更是研究中国古典诗词不可或缺的重要材料。因此，笔者以 CPDB 收录的《全宋词》（第 3 版）作为实证研究的主语料，其正文部分采用 JSON 格式存储，每首词包含词人姓名、正文与词牌名等项内容。

（2）《全宋词典故辞典》（配套语料）。针对《全宋词》而专门编纂的典故辞典有多部，经充分比较后，笔者选用范之麟等主编的《全宋词典故辞典》[2]。该辞典的典故条目按笔画顺序排列，共 10 516 条，每条典故诠释内容包括出典、释义、例句等项，如图 1 所示。中国知网（CNKI）收录了该辞典的电子版全文，笔者将其抓取至本地存储备用。

需要说明的是，目前基于 NLP 技术的文本复用检测、典故自动识别等研究尚未达到实用水平，因此诗词用典分析还只能依靠诗词专家编纂的各种典故辞典来进行。

（3）《全宋词作者索引》（辅助语料）。《全宋词作者索引》附录于《全宋词》正文之后。目前该索引尚未有电子版，需将纸质版扫描，并通过文字识别网站（https://ocr.wdku.net/）以及必要的手工处理后，将其转换为结构化数据存储备用。

3.1.2 语料预处理

语料预处理主要解决语料的结构化、数据纠错补全、噪音去除等问题。

表 1　古诗词用典与学术论文引用的对照比较

比较主题	比较内容	古诗词用典	学术论文引用
用典现象/引文现象	历史起源	始于《诗经》,历史悠久	由学术界参考习惯演变而来,至 19 世纪 50 年代逐渐规范化
	标识符	以典故的典面为主	统一编码体系(顺序编码制或“作者＋出版年”)
	一典多源	有	早期存在科学发现优先权争议现象,后期基本无
	同符异源	有	基本无
	异符同源	有	无
用典动机/引用动机	动机类型	证言、衬言、代名、代言	规范主义理论说 建构主义理论说
	影响因素	文学创作,修辞手段	学术研究需要
用典方式/引用方式	明引与暗用	暗用为主,明引为辅	显式为主,隐式为辅
	单引与叠引	均有,叠引方式更丰富	单引、多引均有
用典行为/引文行为	重复用典	有	有(CountX 方式)
	集句用典	有	无(禁止)
用典位置/引用位置	出现位置	题目＋题序＋正文	正文
用典关系/引用关系	工具依赖	典故辞典、注家注释等	无
	耦合关系	典故耦合关系;典源文献耦合关系	文献耦合关系;耦合关系推广
	同被引关系	典故同被引关系;典源文献同被引关系	文献同被引关系;同被引关系推广
	特殊关系	“一典多源”引发的一对多关系	无
		典源文献因“一典多源”建立的关系	
		檃括词引发的特殊一对一关系	
		注典产生的解释/被解释关系	
用典网络/引文网络	网络构建	构建较困难,且较难延伸	自动构建,较容易
用典失范/引用失范	内容失真	典故误用	引文内容失真(有实质性引文)[44]
	伪造出处	伪注	注而未引(无实质性引文)
	特殊失范行为	注典出错	无
	抄袭行为	“偷诗”行为(具有独特价值)	学术抄袭

一十二都门
【出典】《三辅黄图》卷一《都城十二门》:“长安城东出南头第一门曰霸城门……长安城东出第二门曰清明门……长安城东出北头第一门曰宣平门。……长安城南出东头第一门曰覆盎门……长安城南出第二门曰安门……长安城南出第三门曰西安门。……长安城西出南头第一门曰章城门……长安城西出第二门曰直城门……长安城西出北头第一门曰雍门。……长安城北出东头第一门曰洛城门……长安城北出第二门曰厨城门……长安城北出西头第一门曰横门。”《汉书》卷九九下《王莽传下》:“兵从宣平城门入,民间所谓都门也。”
【释义】都门即都城之门。汉都长安城有十二门,《三辅黄图》记此有“都城十二门”之目。后世因用作咏都城的典故。
【例句】一十二都门,梦想能频。无言桃李几经春。(贺铸《浪淘沙》[一十二都门]525)这里以一十二都门代指宋都开封。

一十二都门	【出典】《三辅黄图》卷一《都城十二门》：“长安城东出南头第一门曰霸城门……长安城东出第二门曰清明门……长安城东出北头第一门曰宣平门。……长安城南出东头第一门曰覆盎门……长安城南出第二门曰安门……长安城南出第三门曰西安门。……长安城西出南头第一门曰章城门……长安城西出第二门曰直城门……长安城西出北头第一门曰雍门。……长安城北出东头第一门曰洛城门……长安城北出第二门曰厨城门……长安城北出西头第一门曰横门。”《汉书》卷九九下《王莽传下》：“兵从宣平城门入，民间所谓都门也。” 【释义】都门即都城之门。汉都长安城有十二门，《三辅黄图》记此有“都城十二门”之目。后世因用作咏都城的典故。 【例句】一十二都门，梦想能频。无言桃李几经春。(贺铸《浪淘沙》[一十二都门]525)这里以一十二都门代指宋都开封。
一丁不识	【出典】《新唐书》卷一二七《张嘉贞传》附《张弘靖传》：“其诟责士皆曰‘反虏’，尝曰：‘天下无事，而辈挽两石弓，不如识一丁字。’军中以气自任，衔之。”宋·吴曾《能改斋漫录》卷五《辨误》：“《唐书·张宏靖传》：‘背挽两石弓，不如识一丁字。’旧史亦同。窦苹《唐书音训》云：‘丁恐当作个。’予尝以窦说虽当，而无所据。偶读孔毅父《续世说》，引宏靖曰：‘汝曹能挽两石弓，不若识一箇字。’乃作此箇字。因知箇误为丁，无可疑者。” 【释义】唐人张弘靖有“能挽两石弓，不如识一个字”语，史书误为“不如识一丁字。”后因以“一丁不识”表示不识字、没文化，也借以喻指没有文化的人。 【例句】君休说是谈非。是则是干支带得来。也要他有个、读书种子;一丁不识，富贵何为。(任翔龙《沁园春·赠谈命许丈》3588)这里以“一丁不识”同“读书种子”对举，说一字不识，没有文化，就算富贵，又有什么用。
一人口插几张匙	【出典】宋·范成大《范石湖集》卷二六《丙午新正书怀十首》其四：“口不两匙休足谷，身能几屐莫言钱。扫除一室空诸有，庞老家人总解禅。”自注：“吴谚云：‘一口不能著两匙。’” 【释义】吴地谚语有“一口不能著两匙”语，谓一张嘴中不能同时放入两个汤匙，比喻人的需求有限，不可贪多。宋词中用为典实。 【例句】①千年田换八百主，一人口插几张匙?休休休，更说甚，是和非。(辛弃疾《最高楼·吾拟乞归，犬子以田产未置止我，赋此骂之》1894)这里化用吴地谚语，意思是说人的享用有限，无需过分追求。用以驳其子“以田产未置止我”的要求。②一人口插几张匙?何用波波劫劫没休时。(刘学箕《虞美人》[寒来暑往何时了]2439)这里用“一人口插几张匙”表示人生岁月无多，享用有限，不必追求。
一力胜九牛	【出典】《列子·仲尼》：“公仪伯以力闻诸侯，堂溪公言之于周宣王，王备礼以聘之。公仪伯至，观形，懦夫也。宣王心惑而疑曰：‘女之力何如?’公仪伯曰：‘臣之力能折春螽之股，堪秋蝉之翼。’王作色曰：‘吾之力能裂犀兕之革，曳九牛之尾，犹憾其弱。女折春螽之股，堪秋蝉之翼，而力闻天下，何也?’” 【释义】周宣王自称：他的力气之大足以拉住九头牛的尾巴，使它们倒退而行。后世常用“一力胜九牛”表现力气超常。 【例句】习懒非痴，觉迷是病，一力那能胜九牛。(赵师侠《沁园春·和伍子严避暑二首》其一2076)这里反用本典，表示生活当顺任性情，听其自然，不要作力不从心的事。
一山红叶供不尽相思句	【出典】唐·范摅《云溪友议》卷下《题红怨》：“明皇代……卢渥舍人应举之岁，偶临御沟，见一红叶，命仆搴来。叶上乃有一绝句。置于巾箱，或呈于同志。及宣宗既省宫人，初下诏，许从百官司吏，独不许贡举人。渥后亦一任范阳，获其退宫人，睹红叶而吁怨久之，曰：‘当时偶题随流，不谓郎君收藏巾箧。’验其书，无不讶焉。诗曰：‘水流何太急，深宫尽日闲。殷勤谢红叶，好去到人间。’”宋·刘斧《青琐高议》卷五《流红记》：“唐僖宗时，有儒士于祐晚步禁衢间。于时万物摇落，悲风素秋，颓阳西倾，羁怀增感。视御沟浮叶，续续而下。祐临流浣手，久之，有一脱叶差大于他叶，远视之若有墨迹载于其上，浮红泛泛，远意绵绵。祐取而视之，果有四句题于其上，其诗曰：‘流水何太急?深宫尽日闲。殷勤谢红叶，好去到人间。’祐得之，蓄于书笥，终日咏味……然莫知何人作而书于叶也。因念御沟水出禁掖，此必宫中美人所作也。……祐自此思念，精神俱耗。……祐终不废思虑，复题二句，书于红叶上云：‘曾闻叶上题红怨，叶上题诗寄阿谁?’置御沟上流水中，俾其流入宫中。……祐后累举不捷，迹颇羁倦，乃依河中贵人韩泳门馆，得钱帛稍稍自给，亦无意进取。久之，韩泳召祐谓之曰：‘帝禁宫人三千余得罪，使各适人，有韩夫人者……吾言之使聘子，何如?’祐避席伏地曰：‘……安敢复望如此!’泳乃令人通媒妁……祐就吉之夕，乐甚。……既而韩氏于祐书笥中见红叶，大惊曰：‘此吾所作之句，君何故得之?’祐以实告。韩氏复曰：‘吾于水中亦得红叶，不知何人作也?’乃开笥取之，乃祐所题之诗，相对惊叹，感泣久之。”参见“一叶怨题”条。 【释义】唐宋笔记小说载：有宫女题诗于红叶，红叶从御沟流出禁苑，被卢渥、于祐得到。日后，二人均与题诗宫女结为夫妇。后世常用作咏男女思恋的典故。 【例句】一山红叶为谁愁，供不尽、相思句。(吕渭老《一落索》[蝉带残声移别树]1127)这里化用红叶题诗典，夸张表现主人公的相思愁怀。
一千龄	【出典】《诗经·鲁颂·閟宫》：“俾尔昌而大，俾尔耆而艾。万有千岁，眉寿无有害。” 【释义】《閟宫》为鲁臣颂美鲁僖公之作。诗中有祝愿鲁僖公活至千岁、万岁之语，后因用作祝愿长寿的典故。 【例句】不须更觅老人星;但愿一年一上一千龄。(李弥逊《虞美人·宜人生日》1059)此词为祝贺生日之作，这里用“一千龄”表示祝愿长寿。

图 1　《全宋词典故辞典》条目示例

(1)《全宋词》语料预处理。主要预处理操作包括:①将 JSON 格式的《全宋词》进行初步结构化处理,获宋词 21 051 首,剔除无名氏词作 1 568 首,得到用于实证分析的宋词 19 483 首。结构化后的基本字段有词 ID、词人、词牌名、正文等。②词人姓名补充纠错。将步骤①中的“词人”字段与《全宋词作者索引》中的词人信息进行对照比较,检查词人姓名出错情况,并利用百度汉语(https://hanyu.baidu.com/)所获词人信

息加以校验。③根据《全宋词作者索引》数据对词人姓名进行消歧，并赋予词人 ID。④对每首宋词的正文内容进行分句处理，并计算每句在该宋词中的绝对位置与相对位置。分句处理用于对宋词用典位置的深入分析。

语料预处理完成后，得到包含如下字段信息的《全宋词》数据表：词 ID、词人姓名、词人 ID、词牌名、词句、句位（n）、句位（%）。

（2）《全宋词典故辞典》语料预处理。观察《全宋词典故辞典》的编纂体例及对典故条目的组织，其预处理操作较为复杂，主要步骤包括：①去除典源列多余内容；②根据辞典中的“参”“见”关系，通过正则表达式提取典面，以获取所对应的典源文献；③分离相似典面。

经预处理后，得到包含典故典面、典故 ID、典源文献等字段的宋词典故数据表。

（3）建立《全宋词》与《全宋词典故辞典》的对应关系表。将预处理后的《全宋词》数据表与《全宋词典故辞典》数据表按典面（文字）进行匹配和联接（匹配方式为正向最长匹配），共获 48 454 条数据备用，每条数据所包含字段为词 ID、词人姓名、词人 ID、词牌名、词句、句位（n）、句位（%）、典故典面、典故 ID、典源文献等，具体实例如表 2 所示（涉及辛弃疾词作 3 首）：

表 2 《全宋词》与《全宋词典故辞典》的对应关系（部分）

词 ID	词人姓名	词人 ID	词牌名	词句	句位/n	句位/%	典面	典故 ID	典源文献
3238	辛弃疾	575	西江月	莫把一分增减	3	50	莫把一分增减	6205	《文选》卷一九战国楚・宋玉《登徒子好色赋》；汉・司马相如《美人赋》
3238	辛弃疾	575	西江月	更上层楼一览	6	100	更上层楼	4550	晋・孙楚《之冯翊祖道》诗；唐・王之涣《登鹳雀楼》诗
3239	辛弃疾	575	西江月	天时地利与人和	2	33	天时地利与人和	5023	《孟子・公孙丑下》
3239	辛弃疾	575	西江月	此日楼台鼎鼐，他时剑履山河	4	67	剑履	10187	《史记》卷五三《萧相国世家》
3239	辛弃疾	575	西江月	都人齐和大风	5	83	大风歌	4342	《史记》卷八《高祖本纪》
3240	辛弃疾	575	生查子	夜夜入清溪，听读离骚去	4	100	离骚	807	《史记》卷八四《屈原列传》；《楚辞》卷一《离骚》东汉・王逸《序》
3240	辛弃疾	575	生查子	青山招不来，偃蹇谁怜汝	1	25	青山偃蹇	8490	宋・苏轼《苏轼诗集》卷七《越州张中舍寿乐堂》诗

（4）《全宋词作者索引》预处理。《全宋词作者索引》的预处理主要是由纸版转电子版。此外，考虑到后续实证分析时便于分时段观察宋词用典的演变情况，还需考虑宋词分期的问题。若只简单分北宋、南宋两阶段，则过于粗略。对此，笔者采用王兆鹏先生提出的六阶段法[46]：①承平时代（1017 – 1067 年）；②变革时代（1068 – 1125 年）；③战乱时代（1110 – 1162 年）；④中兴时代（1163 – 1207 年）；⑤苟安时代（1208 – 1265 年）；⑥亡国时代（1252 – 1310 年），并在《全宋词作者索引》数据表中添加“所处阶段”这一字段，以表明各词人所处的宋词分期。最终的《全宋词作者索引》数据表字段包括词人 ID、词人姓名、所处阶段、备注等，具体实例见表 3。

表 3 《全宋词作者索引》结构化数据表（部分）

词人 ID	作者名	《全宋词》册数	《全宋词》页数	所处阶段	备注
570	王 炎	3	2 392	④中兴时代	婺源
571	张 颁	3	2 393	④中兴时代	–
572	杨冠卿	3	2 403	④中兴时代	–
573	崔敦诗	3	2 410	④中兴时代	–
574	刘清之	3	2 411	④中兴时代	–
575	辛弃疾	3	2 412	④中兴时代	–
576	赵汝愚	3	2 412	④中兴时代	–
577	赵善扛	3	2 551	④中兴时代	–
578	赵善括	3	2 554	④中兴时代	–
579	程 垓	3	2 566	④中兴时代	–
580	虞 俦	3	2 597	④中兴时代	–

3.2 实证研究设计

完整的古诗词用典实证研究主要考虑从以下两个层面展开：分维度分析和整合分析。其中，整合分析需建立在分维度分析基础上，并以典故知识图谱（异构网络）的构建作为支撑。限于篇幅，以下仅对分维度分析加以论述，整合分析部分，因涉及多个用典网络的联接及叠加等复杂处理，考虑到篇幅及主题聚焦等因素，拟在另文中展开。

古诗词用典关系具方向性，因此分维度分析可进一步细分为施典维度（主动方）和典源维度（被动方）两类。

3.2.1 施典维度分析

作为用典关系的主动方，施典维度分析主要围绕"用典词人"和"用典宋词"展开，具体分析内容包括：①用典数量(或频率)，用以衡量词人或宋词用典行为的普遍性，可针对特定词人或宋词，也可分句、分时段或进行整体分析；②用典强度，用以衡量特定宋词中的重复用典情况；③用典位置，用以分析典故在宋词中出现的绝对及相对位置的分布情况(需对宋词正文进行分句处理)；④用典耦合强度，用以衡量两首宋词之间联系的紧密程度。

3.2.2 典源维度分析

作为用典关系的被动方，典源维度分析主要围绕"典故"和"典源文献"(典籍)展开，具体分析内容包括：①被引次数(或频率)，用以衡量某典故或典源文献在宋词中被使用的次数及影响范围，可选择分时段分析或整体分析；②被引强度，用以衡量某典故或典籍在特定宋词中的重复被引现象；③被引位置，用以衡量某典故或典籍在宋词正文中被使用的位置分布及特殊性；④典源文献的同被引强度，通过计算同时引用了某两部典籍的宋词数量，来衡量这两部典籍之间联系的紧密程度。同理，可针对典故的同被引强度进行分析。

3.3 宋词用典的实证分析及结果讨论

依3.1和3.2节所述，笔者对宋词用典问题分别完成了施典维度和典源维度的实证分析，所得主要结果及发现简述如下：

(1)宋词用典是一种普遍现象，选择以宋词为例开展实证研究较为适宜。在全部19 483首宋词中，用典宋词15 974首(占比81.99%)，未用典宋词3 509首(占比18.01%)。对于用典宋词，每首平均用典3.03次，其中约80%用典次数在4次以下，另20%用典次数在[5,16]之间(见图2)。用典次数最高达16次的4首宋词分别是苏轼的《戚氏·玉龟山》、刘克庄的《沁园春·安得奇材》、张元幹的《望海潮·麒麟图画》和陈合的《宝鼎现·神鳌谁断》。能在如此短小篇幅内出现如此频密的典故使用，可见这些词人用典技艺之高强。

此外，分阶段观察发现，单首宋词平均用典次数随时间推移呈现明显的上升趋势(见图2，$R^2=0.83$)。

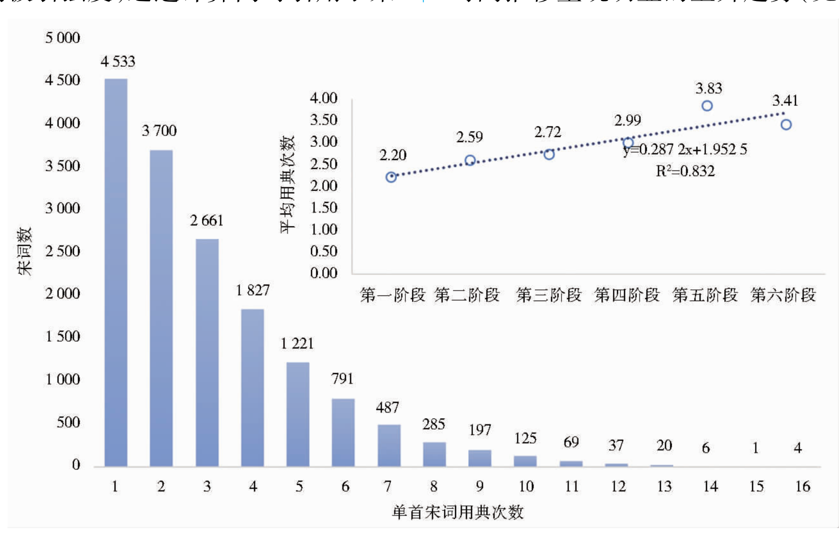


图2 宋词用典频次整体情况及分阶段变化情况

(2)宋词中的用典位置分布具有比较固定的模式，其中绝对位置在第3句达到用典最高点(用典6 935次，占比14.31%)，至第8句后用典次数则急剧下降；相对位置方面，则大体呈现出"三高三低"的"/\/\/"型分布态势，且六阶段的分布与整体分布态势，见图3。

(3)宋词中的重复用典现象比较少见，在正文中多次使用同一典故的宋词仅729首，约占全部用典宋词的4.56%。不过，在重复用典的宋词中发现，有1首宋词使用同一典故达4次，另有1首宋词——洪适的《调笑令·渔父饮时花作荫》，使用同一典故达到了8次之多。

(4)用典宋词的全部典源文献共有2 186部，位列前20位的分别是《文选》《史记》《汉书》《世说新语》《楚辞》《晋书》《后汉书》《诗经》《苏轼诗集》《庄子》《旧唐书》《西京杂记》《南史》《列子》《陶渊明集》《宋书》《三国志》《礼记》《淮南子》和《长恨歌》。其中，《文选》在宋词发展各阶段都影响最大，共被3 855首宋词引用，约占用典宋词的24.13%。另外，分阶段观察可以发现，宋词对这些典籍的使用并不存在文献老化问题，经典文献的影响力可谓是经久不衰。

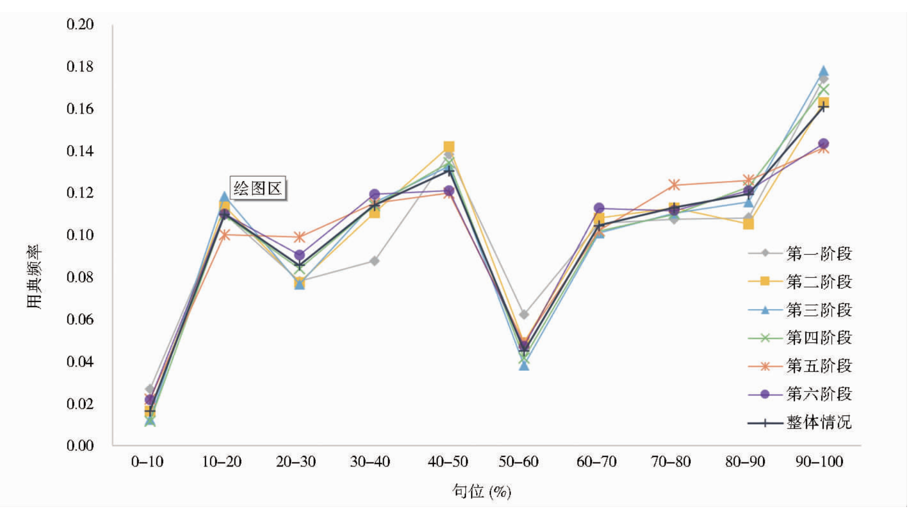


图 3 宋词用典相对位置分布情况

（5）宋词的用典耦合关系比较普遍，但绝大多数耦合强度并不大。在 15 974 首用典宋词中，仅有 169 首宋词与其他宋词之间不存在用典耦合关系；余下的 15 805 首宋词，共形成了 1 608 792 对宋词用典耦合关系，其中 1 589 810 对宋词的典故耦合强度为 1，占比 98.8%。在耦合强度≥2 的宋词对（占比 1.2%）中，发现耦合强度越大，两首宋词间的相似度就越高。一个极端的例子是李清照的《长寿乐・微寒应候》（共 137 字）和苏氏 的《长寿乐・微寒应候》（共 135 字），两者耦合强度最高为 9，字面上仅有 9 字不同。

（6）典源文献间的同被引关系比较普遍，而且联系较紧密。在 2 186 部典源文献中，仅有 152 部典籍没有与其他典籍建立起同被引关系，其余 2 034 部典籍共建立起 53 372 对同被引关系。其中，有 51 057 对典籍的同被引强度＜10（95.66%），1 396 对典籍的同被引强度位于［10，20）区间（2.62%），919 对典籍的同被引强度≥20（1.72%），详情如图 4 所示。图中节点大小表示与该节点具有同被引关系的典源文献数量，边的粗细表示两节点之间的同被引强度。

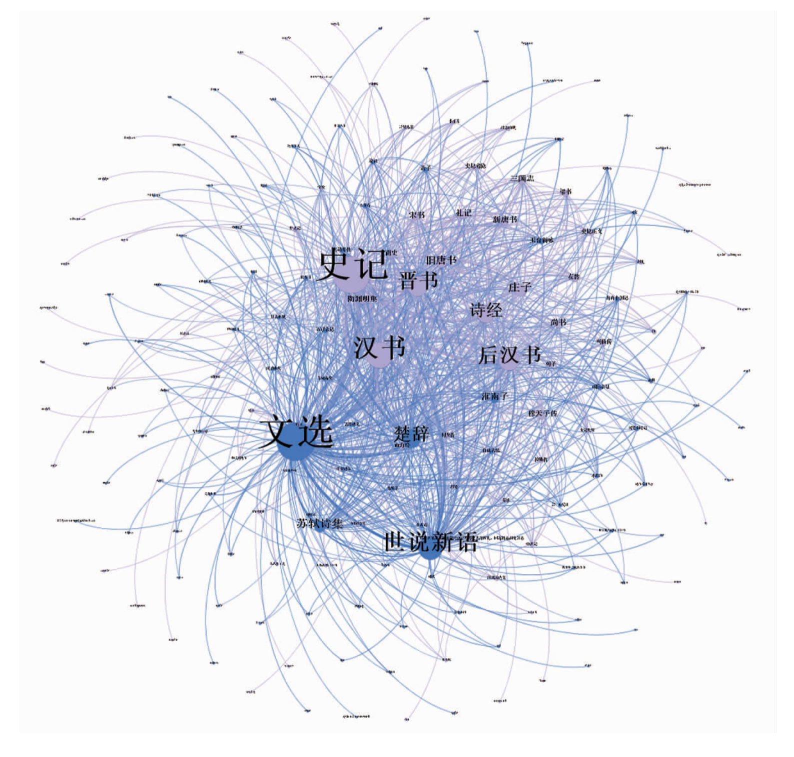


图 4 典源文献同被引关系网络图（同被引强度≥20）

4 结语

笔者主要基于引文分析视角对中国古诗词用典问题展开了较全面的理论梳理,并在此基础上以宋词为例完成了相关的实证分析。初步研究发现,将引文分析法引入到中国古诗词用典研究中,不仅具有理论创新意义,也具备实证研究的可行性,是一次有益的探索和尝试,值得在更多古诗词用典研究中加以借鉴和推广。后续随着研究的不断深入,还可望将其拓展到更多古籍用典问题的分析研究中,这对促进传统人文研究范式由理论阐释到实证解读、由主观判断到客观揭示、由定性分析到定量分析的转变,进而推动数字人文研究领域的创新发展具有很好的方法论意义。

笔者认为,经过半个多世纪的持续发展和完善,引文分析法已进入到全文本计量分析时代,通过与各种NLP技术和可视化工具的日益紧密结合,当前已在引文内容、引文情感、引文语境(上下文)等分析方面不断取得新的研究进展和突破。在此情势下,引言中提及的后两类古诗词用典研究,将在引文分析法中得到有机融合和深化发展。换言之,引文分析法有望成为与传统人文范式并行发展、相互补充的古诗词用典研究的一种新范式,并促进古诗词用典研究逐步形成兼具"理论＋实证""主观＋客观""定性＋定量"的新态势。

最后要强调的是,在本文研究过程中,笔者发现基于引文分析法开展古诗词用典问题的量化研究,面临的最大挑战和研究难点之一是如何对典故进行准确识别和提取。究其原因,主要有以下3个:①诗词用典本身属文学创作范畴,更多体现主观情感和个人偏好,典故理解及其意象很难形成比较客观和一致性的认知;②在涵盖诗词用典在内的更多古籍用典行为中,普遍存在"一典多源""同符异源""异符同源""典中典"等特殊现象,不要说自动识别,即使是专家注典或进行人工判断,也难免出错;③现有的文本复用检测、典故自动识别等相关算法还处于探索阶段,并不足以对用典研究形成有力的技术支撑。现阶段,古诗词用典的引文分析还只能优先选择已有前人编纂的配套典故辞典的诗词语料来进行,期待未来能在典故自动识别方面有所突破,以不断拓展诗词用典的研究范围和研究深度,并在古诗词用典知识图谱构建及可视化展示、多维度查询及计量分析等方面形成更多实用性的成果。

参考文献:

[1] 谌东飚. 唐前诗歌用典考察[J]. 长沙水电师院社会科学学报, 1996(4):57－62.
[2] 范之麟, 刘扬忠, 徐燕. 全宋词典故辞典[M]. 武汉:湖北辞书出版社, 2001:1.
[3] 林岗. 论引诗[J]. 文艺理论研究, 2007(4):43－49.
[4] 王思豪. 一个被遮蔽的语体结构选择现象——论汉赋用《诗》"《诗》曰"的隐去[J]. 文学遗产, 2013(4):55－61.
[5] 李鹏飞. 中古诗歌用典美学研究[M]. 武汉:武汉大学出版社, 2016:29.
[6] 马草. 诗性功能与文学自觉——魏晋南北朝文学用典的审美原理考察[J]. 哈尔滨工业大学学报(社会科学版), 2018, 20(6):75－81.
[7] 胡大雷. 六朝诗歌用典论——兼论"诗言志"与集体无意识[J]. 文学评论, 2014(5):154－160.
[8] 秦红丽. 晚唐诗歌用典艺术研究[D]. 南京:南京师范大学,2019.
[9] 董朝刚. 也论宋诗特点——关于"以文字为诗,以议论为诗,以才学为诗"的若干思考[J]. 唐山师范学院学报, 2004(4):10－12,61.
[10] 刘颖. 文字、才学、议论——关于苏黄"以文字为诗,以才学为诗,以议论为诗"的若干思考[J]. 安徽文学(下半月), 2008(5):56－57.
[11] 莫砺锋. 黄庭坚"夺胎换骨"辨[J]. 中国社会科学, 1983(5):187－198.
[12] 陈福升. 论词中用典与唐宋词的发展[J]. 内蒙古社会科学(汉文版), 2003(S1):106－107.
[13] 李定广. 论北宋词与晚唐诗的近亲关系——兼论正确解读宋词化用唐诗现象的文化涵义[J]. 求索, 2006(11):170－172,187.
[14] 吴承学. 论宋代檃括词[J]. 文学遗产, 2000(4):10.
[15] 郑园. 论东坡檃括词[J]. 文学遗产, 2006(3):147－149.
[16] 范进军. 试论稼轩词用典的方式[J]. 船山学刊, 2006(1):111－114.
[17] 张岳伦. 中国古典诗词用典简论[J]. 湖南人文科技学院学报, 2007(1):138－140.
[18] Chinses－Poetry 中国诗歌古典文集数据库(CPDB)[EB/OL]. [2022－06－13]. https://github.com/chinese－poetry.
[19] GARFIELD E. Can citation indexing be automated? [C]// STEVENS M E, GIULIANO V E, HEILPRIN L B. Statistical association methods for mechanized documentation: symposium proceedings. Washington DC: The National Bureau of Standards, 1965:189－192.
[20] WEINSTOCK M. Citation indexes[M]//KENT A. Encyclopedia of library and information science: Vol.5. New York: Marcel Dekker,1971:16－40.
[21] THORNE F C. The citation index: another case of spurious validity [J]. Journal of clinical psychology, 1977, 33(4): 1157－1161.
[22] BROOKS T A. Private acts and public objects: an investigation of citer motivations[J]. Journal of the American Society for Information Science, 1985, 36(4):223－229.
[23] BORNMANN L, DANIEL H. What do citation counts measure? a review of studies on citing behavior[J]. Journal of documentation, 2008, 64 (1) : 45－80.
[24] 罗积勇. 用典研究[M]. 武汉:武汉大学出版社, 2005.
[25] 舒非, 丰鹂萱, 邱均平,等. 基于我国古籍引经据典现象的引文分析研究[J]. 情报学报, 2021, 40(12): 1338－1346.
[26] 朱栋. 唐代试律诗用典研究[D]. 武汉:武汉大学, 2013.

[27] 陈望道. 修辞学发凡[M]. 上海:上海人民出版社, 1976:97.
[28] 范之麟, 吴庚舜. 全唐诗典故辞典[M]. 武汉:湖北辞书出版社, 1989.
[29] 彭庆生, 曲令启. 诗词典故词典[M]. 太原:书海出版社, 1990.
[30] 赵应铎. 中国典故大辞典(辞书版)[M]. 上海:上海辞书出版社, 2014.
[31] 姚蕊. 基于典故领域本体的诗词用典自动分析系统[J]. 软件导刊, 2011,10(6):80 - 82.
[32] 搜韵[EB/OL]. [2022 - 08 - 07]. https://sou - yun. cn/index. aspx.
[33] TANG X, LIANG S, ZHENG J, et al. Automatic recognition of allusions in Tang poetry based on BERT[C]//2019 international conference on Asian Language Processing (IALP). Piscataway: IEEE, 2019.
[34] 张明华. 集句诗的发展及其特点[J]. 南京师范大学文学院学报, 2006(4):24 - 29.
[35] 刘成国. "荆公体"别解[J]. 文学遗产, 2006(4):149 - 151.
[36] 左汉林, 高付军. 宋代集杜诗发展的三个阶段及其特征[J]. 渭南师范学院学报, 2012, 27(11):88 - 92.
[37] 唐圭璋, 缪钺, 叶嘉莹, 等. 唐宋词鉴赏词典(唐・五代・北宋)[M]. 上海:上海辞书出版社, 1988:657.
[38] 文天祥. 文山先生全集[M]. 北京:北京市中国书店, 1985:387.
[39] 皎然. 诗式校注[M]. 周维德, 校注. 杭州:浙江古籍出版社, 1993:31.
[40] 张明华. 略论集句诗的文献价值及其相对性[J]. 文学遗产, 2012(2):91 - 100.
[41] 卞仁海. 李善的征引式注释[J]. 信阳师范学院学报(哲学社会科学版), 2005(3):78 - 82.
[42] 萧统.《文选》[M]. 李善, 注. 北京:中华书局, 1977:329.
[43] 王乙珈. 用典与"注典"——宋代杜甫诗注中的《世说新语》及其学术意义[J]. 北京社会科学, 2021(3):13 - 24.
[44] 刘应竹. 学术论文中的引文失范问题刍议[J]. 编辑学报, 2014, 26(1):7 - 9.
[45] 唐圭璋. 全宋词[M]. 北京:中华书局, 1965.
[46] 王兆鹏. 论宋词的发展历程[J]. 暨南学报(哲学社会科学), 2000(6):14 - 23,29.

作者贡献说明:
蔡竹娟:论文选题及写作,数据处理及实证分析;
赵丹群:选题指导,论文写作及后期修改。

Research on Allusions Usage in Ancient Chinese Poetry from the Perspective of Citation Analysis

Cai Zhujuan Zhao Danqun
Department of Information Management, Peking University, Beijing 100871

Abstract: [Purpose/Significance] From the perspective of citation analysis (CA), the paper theoretically analyzes and empirically studies the use of allusions in ancient Chinese poetry in order to promote the transformation and expansion of paradigm from traditional humanistic research to data - driven quantitative research in the era of big data, which is helpful in integration and innovative development of ancient Chinese poetry allusion studies and bibliometric methods. [Method/Process] Firstly, the paper gave a completely comparative discussion focusing on their similarities and differences between the allusion behaviors (phenomenons) in ancient Chinese poetry and the citation behaviors (phenomenons) among academic papers from the following three aspects: ① ancient Chinese poetry allusion phenomenon and its emergence & development; ② motivations and behaviors while using allusions; ③ allusion relationships and allusion networks. Secondly, a quantitive empirical study about allusions using based on *The Complete Song Ci* & *The Allusions' Dictionary of Song Ci* was carried out in two dimensions: the literatures of using allusions and the literatures of producing allusions. [Result/Conclusion] The phenomenon of using allusions in ancient Chinese poetry has a long history along with rich topics to be studied, the innovation and transformation of research paradigm are needed urgently at present stage. Through theoretical analysis and empirical results, it can be seen that the citation analysis method is helpful to carry out more in - depth exploration on a large scale of ancient Chinese poetry corpus. The biggest difficulty in current research lies in automatic recognition and extraction of allusions. In the future, it is necessary to strengthen the close combination with NLP technology, and strive to put forward more effective automatic solutions and quantitative analysis strategies in allusion identification and extraction, motivation and behavior classification of using allusions, allusion network and allusion knowledge graph construction, so as to effectively promote the transformation of traditional humanistic research paradigm from theoretical analysis to empirical interpretation, from subjective judgment to objective disclosure, from qualitative analysis to quantitative analysis, and produce innovative results on more Digital Humanities (DH) research topics.

Keywords: citation analysis (CA) ancient Chinese poetry allusion(s) motivation of using allusions allusion behaviors